Thresholding Approach based on GPU for Facial Expression Recognition

Jesús García-Ramírez[1], J. Arturo Olvera-López[1], Ivan Olmos-Pineda[1], Georgina Flores-Becerra[2], Adolfo Aguilar-Rico[2]

[1] Benemérita Universidad Autónoma de Puebla, Faculty of Computer Science, Puebla, México
gr_jesus@outlook.com,{aolvera,iolmos}@cs.buap.mx
[2]Instituto Tecnológico de Puebla, Puebla, México
{kremhilda,adolforico2}@gmail.com

Abstract. Facial expression recognition and its applications are currently research topics in computer science. When analyzing face images, it is necessary to apply an algorithm for pre-processing them in order to extract features from each facial expression. In this paper related to our initial research phase, we propose variations and results from algorithms for binarizing and thresholding face expression images. In addition, we present earlier results about the parallel implementation of our proposed algorithms.

Keywords. Facial Expression Recognition, Image pre-processing, Segmentation, Thresholding.

1 Introduction

In recent years, several applications have been developed about information processing, in special those related to digital images, such as facial expression recognition; some applications in this field are detection of mental disorders, detecting whether a person is lying in an interview, detection of emotions, among others. When processing digital images, it is very common to find regions of interest (ROI), i.e. extracting from the whole image those regions to be analyzed. For extracting ROI's there exist several methods, for example those based on analysis of image textures, and some other based on locating main points related to eyebrows, nose and mouth.

The process of facial expression recognition starts with a pre-processing stage usually related to tasks such as: smoothing, scaling, converting to gray scale, among others. After the pre-processing step, a process for feature extraction is applied; first, ROI's are extracted from the whole image using techniques where the background and non-background are detected. After that, a second stage is performed, where the relevant features are extracted for expressions analysis, such as eyes, eyebrows, mouth and cheeks.

Commonly the face expression analysis is carried out over images represented in RGB (Red, Green, and Blue) color model, however, other color models can provide relevant information about border, luminance or texture. For this reason, it is important to analyze the information represented by other color spaces like HSV or YUV.

When processing digital images every pixel is taking into account and depending of the image size, this process could be computationally expensive. Due the processing of digital images requires many computation resources, in special when processing video, for example, some video devices capture 1500 fps (frames per second, i.e. 1500 images per second), then the runtime and the amount of resources are huge.

Alternatives for dealing with these problems are the parallel approaches such as the architectures provided by Graphic Processing Units (GPU) in personal computers. Initially these devices were used only for graphic processing, but nowadays they can be used for parallel programing taking advantage of the processing cores into them.

In this paper we present results about our early methodology stage for pre-processing digital images for facial expressions recognition, we present an approach for pre-processing face images based on both sequential and GPU approaches.

The paper is organized as follows: Section 2 describes the problem to solve, Section 3 show the methods and the related works of facial expression recognition, Section 4 presents the proposed methodology for threshold and pre-processing, then experiments results are shown; finally, in Section 5 conclusions and future work are discussed.

2 Problem to Solve

The human-computer interaction has as main objective, the natural interaction with humans through a hybrid computer system (hardware and software), it means that the interaction must be like humans communicate among them. The human communication with each other is commonly done via gestures, speech or written sentences. But there exists another way for communicating among humans, which is related to the facial expression involved in most of the oral communications, the facial expressions are found in 55 percent of communicated messages, 38 percent and 7 percent related to voice intonation and spoken words respectively [1].

The communication among humans is mainly characterized by the fact of that, the humans can express emotions during a conversation. On the other hand, the computers are not able neither show nor understand emotions, for this reason a natural human-computer interaction is computationally difficult.

Nowadays, the basic approach about detection of facial expressions consists in analyzing the interaction between the computer and the human by a digital capture device and then a computer system will classify the emotion, in this process the number of images to process will be grow directly according the amount of images captured per second. This process could be computationally expensive, then an alternative is to develop facial expression recognition algorithms by parallel approaches, reducing execution time and resources.

The process that several researchers follow in this kind of approaches is shown in Fig. 1. The input of the system is an image or multiple images (video), and then the pre-processing image step is followed; this process can implement image filters, border detection and thresholding methods to get the descriptive face information. The next stage of the system is the feature extraction, encoding the image information to numeric values, then the system uses a model previously constructed (by a learner or classifier) to classify person's facial expression in the image.

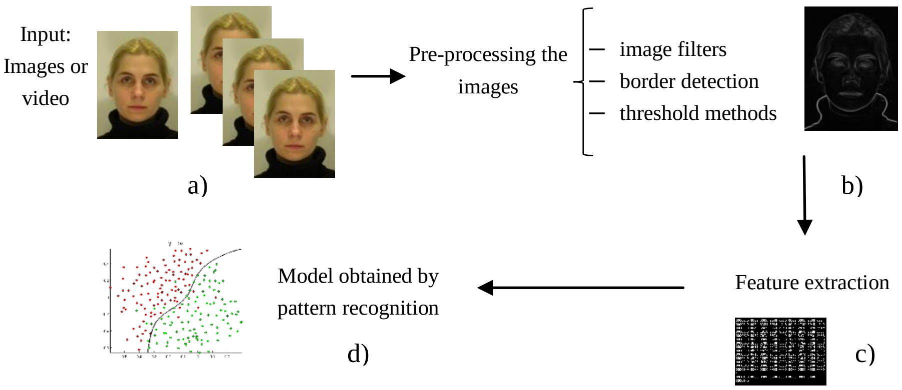

Fig. 1. Process commonly used for facial expression recognition, a) input images, face images taken from [16], b) pre-processing stage, c) feature extraction, d) model created by the classifier from information in c)

3 Related Works

The facial expression recognition is a research area since 1970 when Ekman presented two important contributions. The first one presents the six universal facial expressions: happiness, anger, sadness, disgust, surprise, and fear, several approaches about facial expression recognition find these emotions [2]. The second one is the Facial Expression Coding System (FACS), this system describes the movements in the face and also gives a time window to detect these movements. Action Units (AU), are smallest visually discriminable movements with some qualification in the FACS, this codification system specifies 9 AUs in the upper face, and 18 in the lower face. In addition, there are 14 head positions and movements, 9 eye position and movements, 5 miscellaneous actions units, 9 action descriptors, 9 gross behavior and 5 visible codes [3].

As it is shown in Fig.1, the first stage of the face expression system consists of pre-processing the images, in this process the system prepares the images to extract the principal features. Before implement pre-processing, the images can be transformed to another color spaces like reported in [4] where it is shown that other color space provide information different from that described by the RGB space. They show experiments with HSV, RGB, and YUV color models, as show in the Fig.2 the channels SV show more information about edges than RGB model about the image, and H channel do not give any relevant information of the image.

a) b) c) d)

Fig. 2. a) Image represented in the RGB model b) the channel H, c) channel S d) channel V

Other pre-processing method is show in [1], this method is based on edge detection and a comparison among four border detection algorithms is reported, the algorithms compared are: Robert, Sobel, Laplace and Canny, having this last the best detection performance among the four algorithms.

Bourel F. et al. [18] report that 12 points are localized in the face of the eyebrows, nose, and mouth, then according to the distance between some of them, one of three possible states (increase, stable or decrease) is assigned for detecting the expression. Some methods pre-process the images using a threshold, which takes into account the pixels that have a higher value than a threshold. Other approaches implements models in 3D [6, 7], combine an extract Local Binary Pattern (LBP) for face localization, they use a pseudo 3D model, combining this methods to find the ROI's and extract the features. The local ROI's that several works take into account are: eyebrows, eyes, nose and mouth [8].

The feature extraction, registers the information of pre-processed images; there exist three method for information registration [10]: whole face (the face as hole entity, or register locally information), only parts of the face (like eyes, eyebrows, mouths, etc.) and edge of face (find a shape in the image). An example for registering the whole face is presented in [9], this approach pre-process the images obtaining only the face information, then the image is splited and a Gabor filter is applied, Gabor filter is used in digital images in order to analyze the information in the frequency space [19]. Other works use 3D images and find the main points of the face, then the distance between these points is registered; other approaches register the LBP from histogram of a 3D image [20, 21].

For a local registration, the register information is that related to Motion Units (MU), the MU are simply movements of face's muscles, such as movements for the eyes, eyebrows, lips, and cheeks[11].

The final stage in facial expression recognition is the classification, it uses the information from the feature extraction stage to find a model for classify, such models, are extracted by classifiers to predict or classify [22]. The classifiers use a dataset for extracting models i.e. the training set, when the dataset is previously categorized, the learning of classifier is called supervised learning, otherwise it is called unsupervised learning. Several classifiers have been applied for face expression recognition, in [11] N. Sebe et al. present a comparison among the classifiers: Naïve Bayes, Bayesian Networks, Decision trees, Nearest Neighbors. Among these classifiers, the best performance was obtained by decision tree MC4. In [12, 13] an approach that implements a 3D model and a classification with Hidden Markov Models (HMM) is presented, the performance are good, but only three expressions are found (Happy, Sad, Surprise), most of the approaches implements HMM as show in [10, 14]. Other approaches like [15] use Neural Network for the classification stage, and [9] implements Nearest Neighbors classifiers.

4 Proposed Methodology and Results

In this work we present the results of the initial stage of a facial expression recognition system, we show a comparison among different technics for border detection and some variations from them. Then we propose a method to find a threshold for binarizing an image, a comparison among Isodata and Otsu threshold methods are reported. Additionally we analyze face images in different color spaces; in our experiments, we used RGB images from the MMI Database [16] and the Cohn Kanade+ Database [17].

In order to find the face and the AU regions (eyes, eyebrows, nose, cheeks, mouth), border extraction is used for detecting these regions. We applied the following border detection algorithms: mixed vertical/horizontal, gradient, Laplace, Sobel filters [1], and some variations of these.

Some obtained results are reported in Fig. 3. In order to compare these results, we computed the Structural component Coefficient (SC) which quantifies the quality (related to the original image) of the obtained image; the higher SC value the best quality in the image and $0 \leq SC \leq 1$. The expression for computing SC is described in equation (1), where $f(i, j)$ denotes the intensity value at position (i, j) in the original image whereas $f'(i, j)$ is related to the values in the resultant image.

$$SC = \frac{\sum_{j=1}^{M}\sum_{k=1}^{N}[f(j,k)]^2}{\sum_{j=1}^{M}\sum_{k=1}^{N}[f'(j,k)]^2} \qquad (1)$$

The Sobel image have the highest SC value, but the image result to apply the Sobel filter contains noise that make difficult to process it, this problem is because the border transition in the original images are so faint. To solve this problem, a variation of the edge detection filters is experimented in this work; the traditional filters take into account the neighbors pixels and consider a convolutional matrix of 3x3 pixels. Then the variation we implement is to consider convolutional matrix with higher dimension depending on a value u, and then we use a matrix of $(2u+1)x(2u+1)$ and consider only the border pixels of the convolutional matrix.

The Fig. 4 shows the performance of the border filters (vertical/horizontal, gradient, Laplace, and Sobel) using u=5 and their SC values, the obtained result is better than the traditional filters. The main disadvantage is that for the higher value of u the higher amount of noise in the image is obtained, but this variation results better because the borders in the image are very soft, and other issue is that for the higher u value, the thicker border in the image is obtained. Again, the Sobel filter have the highest correlation coefficient, but the amount of noise in the image is higher than in traditional filter.

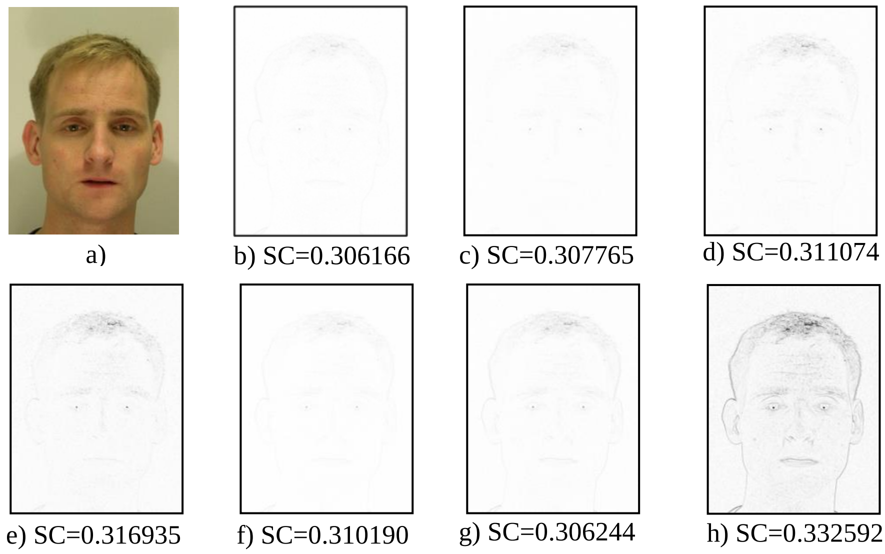

Fig. 3. Results of the border detection filter and their corresponding SC values, a) image in RGB color space, b) mixed vertical/horizontal filter, c) Laplace edge detection, d) Laplace 45 degrees, e) full Laplace filter, f) Gradient 45 degrees, g) Gradient filter, h) Sobel filter

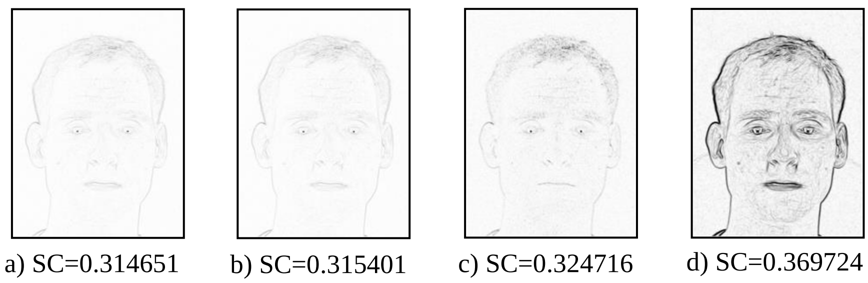

Fig. 4. Results obtained using *u*=5, a) horizontal/vertical, b) Gradient, c) Laplace, d) Sobel

To find the edge information in an image it is common binarizing in the image. This process consists of assigning a value to each pixel either 0 or 255 from a threshold. We propose a Thresholding Method based on Average (TMA), which is shown in Fig. 5. In this method, first any edge detection algorithm is applied over a gray scale image (I_B), then the frequencies histogram and the mean (*Avg*) from I_B are computed, the threshold *t* computed by TMA uses these frequencies; *t* corresponds to the maximum difference between two consecutives restricted to only considering frequencies higher than *Avg*. Mainly, *t* is the higher value among significant transitions.

We make a comparison against Otsu and Isodata algorithm, the results of applying the three methods is shown in Fig. 6, note that the TMA algorithm give the best edge information, less noise, and the best edge information among Otsu and Isodata.

TMA (Threshold Method based on Average)
Input: I (Image in RGB Format), Output: t (threshold)

01. $I_G \longleftarrow$ Transform I to gray scale format
02. $I_B \longleftarrow$ Apply any edge detection algorithm
03. $F_{IB} \longleftarrow$ Compute the histogram of I_B
04. $Avg \longleftarrow$ Compute the mean of F_{IB}
05. For $i = 0$ to 255 do
06. If $F_{IB}(i) \leq$ avg then $F_{BI}(i) = 0$
07. $t \longleftarrow \forall_{0<i<254}: \max\{F_{IB}(i) - F_{IB}(i+1)\}$
08. Return t

Fig. 5. TMA Algorithm for finding a binarizing threshold in a RGB image

a) b) c)

Fig. 6. Thresholding results obtained by a) Isodata, b) Otsu c) TMA algorithm

Another way of pre-processing face images consists of transforming the images into another color space. In the Fig. 2 the channels of the HSV color space are shown, then we can apply two other color spaces such as YUV and L*a*b*. In Fig. 7 the results of converting RGB to this models are depicted, channels Y and L* do not give relevant information, these channels show a gray scale image because they only consider information of luminance.

a) b) c) d)

Fig. 7. Face images in different color models a) U channel of YUV color space b) the V channel of YUV color space, c) a* channel of L*a*b* color model d) b* channel of L*a*b*

Before the system starts to implement a threshold algorithm or a border detection, the system needs to locate the face in the image. A transformation into other color space can help us for locating the face and interest points in the image, the best among our experimental results is the b* channel of the L*a*b* color space. As shown in the Fig. 8a, the marks of the eyes, mouth, nose and the border of the person in the image are sharped. Then a border detection method like Laplace can be applied to find the border of the face and the interest points as shown in Fig. 8b.

Fig. 8. Face location a) transformation of the image in figure 3a to b* channel and the implementation of Otsu threshold algorithm, b) Laplace border detection

The previous implementation was applied over the global image (i.e. using a global binarizing threshold), but there exists other kind of implementation for only a local region in the image, this implementation can be applied over sub-regions of at most 32x32 pixels (i.e. using local binarizing threshold) and this can be implemented by a parallel approach. Applying the edge detection and the threshold methods over a local region better results can be obtained, because the noise in a region of the image are less than in the whole image, as can be seen in the Fig. 9 we applied the Otsu method over three regions of an image, the local edge detection is better than the global implementation.

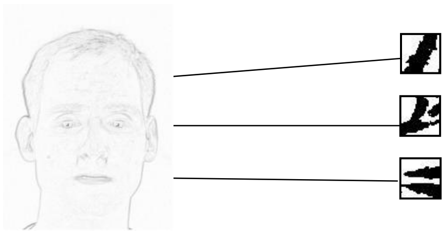

Fig. 9. Results obtained after applying local thresholding over 32x32 regions

As we mentioned before, processing face images could be expensive, in the following paragraphs, we show the results from our parallel implementation of TMA algorithm, for the parallel approach we consider splitting the image in regions and applying a local thresholding to each region in the image.

In our parallel approach we use gray scale images, we implement the TMA parallel. The implementation was applied in a GPU Nvidia with 192 cores with a clock base of 797 MHz and 2 GB in RAM, then we make test using images of different size obtained by image scalling.

In Fig. 10 we show the performance of the parallel approach TMA, note that as the higher size of the image, the higher runtime is reported, but as higher number of cores less runtime is obtained. This experiment let us notice the usefulness of the cores processing when facing to higher dimensionality images. But if we apply this approach in a GPU with higher clock speed, or higher number of cores it can be more efficient than a sequential approach, as we mentioned the image processing is computationally expensive and more when processing video, because not only one image is processed. Then the time of processing frames of video in a sequential approach will increase.

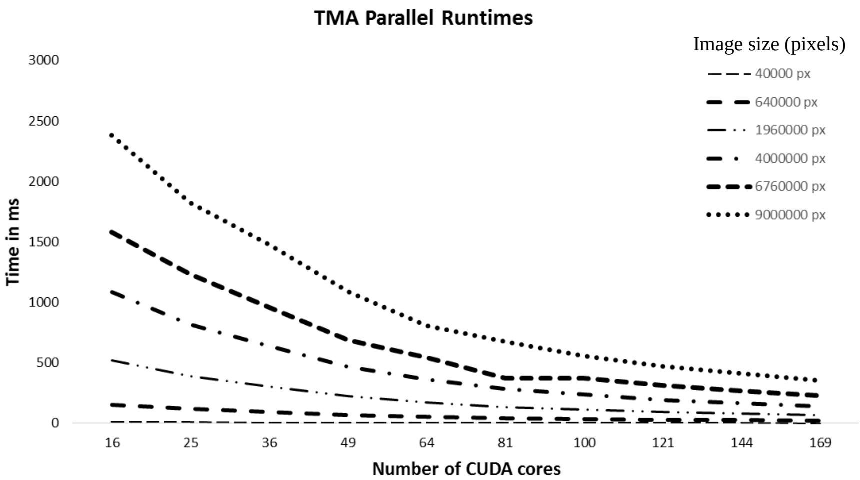

Fig. 10. Runtime of parallel approach of TMA algorithm in milliseconds with different number of cores

5 Conclusions

In order to find points of interest in facial expression recognition system, border detection and thresholding algorithms are used, we propose variations for thresholding and border detection, which have better performance than the traditional edge detection filters, this variation can be used even when the images has soft transitions in borders.

Finally, we will tested our GPU parallel implementation obtaining lower runtimes according to the increase of the number of cores, nevertheless, the load time have an impact in the global time, which makes useful the GPU approach when facing to problems requiring high sequential runtimes.

As a future work we continue in the second stage related to feature extraction from segmented images and then we will propose approaches for classifying facial expressions. In addition, we will construct a facial expression database capturing images from both visible and thermal (infrared) ranges, which will be used for testing our future approaches.

Acknowledgements

This work was partially supported by the CONACyT Mastering Scholarship 701191, the PRODEP network "Desarrollo de tecnologias de la información y comunicación" and projects VIEP-BUAP: OLLJ-ING16-I, OLPI-ING16-I.

References

1. Xiaoming C., Wushan, C.: Facial expression recognition based on edge detection. International Journal of Computer Science & Engineering Survey6(2), 1–9 (2015)
2. Ekman P.: Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique. Psychological Bulletin 115(2), 268–287 (1994)
3. Cohn J., Ambadar Z., Ekman, P.: Observer-based measurement of facial expression with the Facial Action Coding System. In: J. A. Coan & J. J. B. Allen (Eds.), The handbook of emotion elicitation and assessment. New York, pp. 203–221. (2007)
4. Torres L., Reutter J., Lorente, L.: The importance of the color information in face recognition. In: Proceedings of the International Conference on Image Processing, (ICIP 99), Vol. 3, pp. 627–631 (1999)
5. Chibelushi C., Bourel, F.: Facial Expression Recognition: A Brief Tutorial Overview. In: On-Line Compendium of computer Vision, Vol. 9, pp.1–5 (2002)
6. Huiquin Z., Osamu, Y.: A Fusion Approach for Facial Expression Using Local Binary Pattern and a Pseudo 3D Face Model. In: Proceedings of International Conference on Software Engineering, Artificial Intelligence, Networking Parallel/Distributed Computing, pp. 121–126 (2013)
7. Essa I., Pentland, A.: Coding Analysis, Interpretation, and recognition of facial expression. IEEE Transaction on Pattern Analysis and Machine Intelligence 19(7), 757–763 (1997)
8. Yongqiang L., Shangfei W., Yongping Z., Qiang, J.: Simultaneous Facial Feature Tracking and Facial Expression Recognition. IEEE Transactions on Image Processing 22(7), 2559–2573 (2013)
9. Wenfei G., Cheng X., Venkatesh, Y., Dong H., Han L.: Facial expression recognition using radial encoding of local Gabor Features and classifier synthesis. Pattern Recognition 45(1), 80–91 (2012)
10. Sariyanidi E., Gunes H., Cavallaro, A.: Automatic Analysis of Facial Affect: A survey of Registration, Representation, and Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 37(6), 1113–1133 (2015)
11. Cohen I., Sebe N., Gozman F., Cirelo M., Huang, T.: Learning bayesian network classifiers for facial expression recognition both labeled and unlabeled data. In: Proceedings of Computer Vision and Pattern Recognition (CVPR 2003), pp. 595–601 (2003)
12. Le V., Tang H., and Huang, T.: Expression recognition from 3D dynamic faces using robust spatio-temporal shape features. In: Proceedings of Face and Gesture Recognition (FG 2011), pp. 414–421 (2011)
13. Sandbach G., Zafeiriou S., Pantic M., Rueckert, D.: A dynamic approach to the recognition of 3D facial expressions and their temporal models. In: Proceedings of Face and Gesture Recognition (FG 2011), pp. 406–413 (2011)

14. Corneau C., Oliu M., Cohn J., Escalera, S.: Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-related Applicants. IEEE Transactions on Pattern Analysis and Machine Intelligence 99, 1–20 (2015)
15. Tian Y., Kanade T., Cohn, J.: Recognizing Action Units for Facial Expression Analysis. Transactions on Pattern Analysis and Machine Intelligence 232(2), 97–115 (2001)
16. Pantic M., Valstar M., Rademarker R., Maat, L.: Web-based database for facial expression Analysis. In: Proceedings of International Conference on Multimedia and Expo, pp. 5–10 (2005)
17. Kanade, T., Cohn, J., & Tian, Y.: Comprehensive database for facial expression analysis. In: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG 2000), pp. 46–53 (2000)
18. Bourel F., Chibelushi C., Low, A.: Robust Facial Expression Recognition Using a State-Based Model of Spatially-Localized Facial Dynamics. In: Proceedings of the Fifth IEEE International Conference Automatic Face and Gesture Recognition, pp. 106–111 (2002)
19. Littlewort G., Whitehill J., Wu T., Fasel I., Frank M., Movellan J., Bartlett, M.: The computer expression recognition toolbox (CERT). In: Proceedings of Face and Gesture Recognition (FG 2011), pp. 298–305 (2011)
20. Tang H., Huang, T.: 3D facial expression recognition based on automatically selected features. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR 2008), pp. 1–8 (2008)
21. Sandbach G., Zafeiriou S., Pantic, M.: Local normal binary patterns for 3D facial action unit detection. In: proceedings of the International Conference of Image Processing (ICIP 2012), pp. 1813–1816 (2012)
22. Jiawei H., Micheline K. Jian, P.: Data Mining, Concepts and Techniques, Elsevier Inc., (2012)